



Atty. Dkt. No. 017346-0177

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant: Nicolas TSINGOS et al.  
Title: PERFECTED DEVICE AND METHOD FOR THE  
SPATIALIZATION OF SOUND  
Appl. No.: 10/748,125  
Filing Date: 12/31/2003  
Examiner: Unassigned  
Art Unit: 2641

**CLAIM FOR CONVENTION PRIORITY**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

The benefit of the filing date of the following prior foreign application filed in the following foreign country is hereby requested, and the right of priority provided in 35 U.S.C. § 119 is hereby claimed.

In support of this claim, filed herewith is a certified copy of said original foreign application:

- French Patent Application No. 0313875 filed 11/26/2003.

Respectfully submitted,

Date: May 12, 2004

By Brian J. McNamara

FOLEY & LARDNER LLP  
Customer Number: 22428  
Telephone: (202) 672-5416  
Facsimile: (202) 672-5399

Brian J. McNamara  
Attorney for Applicant  
Registration No. 32,789



**THIS PAGE BLANK (USPTO)**



---

# BREVET D'INVENTION

---

## CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

### COPIE OFFICIELLE

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le .....

Pour le Directeur général de l'Institut  
national de la propriété industrielle  
Le Chef du Département des brevets

Martine PLANCHE

INSTITUT  
NATIONAL DE  
LA PROPRIÉTÉ  
INDUSTRIELLE

SIEGE  
26 bis, rue de Saint Petersburg  
75800 PARIS cedex 08  
Téléphone : 33 (0)1 53 04 53 04  
Télécopie : 33 (0)1 53 04 45 23  
www.inpi.fr

**THIS PAGE BLANK (USPTO)**



26 bis, rue de Saint Pétersbourg  
75800 Paris Cedex 08  
Téléphone : 33 (1) 53 04 53 04 Télécopie : 33 (1) 42 94 86 54

# BREVET D'INVENTION CERTIFICAT D'UTILITÉ

Code de la propriété intellectuelle - Livre VI



## REQUÊTE EN DÉLIVRANCE page 1/2



Cet imprimé est à remplir lisiblement à l'encre noire

DB 540 e W / 210502

<b>REMISE DES PIÈCES</b> DATE <b>26 NOV 2003</b> LIEU <b>75 INPI PARIS 34 SP</b> N° D'ENREGISTREMENT <b>0313875</b> NATIONAL ATTRIBUÉ PAR L'INPI DATE DE DÉPÔT ATTRIBUÉE <b>26 NOV. 2003</b> PAR L'INPI		<b>1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE</b> À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE CABINET NETTER 36 avenue Hoche 75008 PARIS	
<b>Vos références pour ce dossier</b> (facultatif) INRIA Aff. 72 (120905)			
<b>Confirmation d'un dépôt par télécopie</b>		<input type="checkbox"/> N° attribué par l'INPI à la télécopie	
<b>2 NATURE DE LA DEMANDE</b>		<b>Cochez l'une des 4 cases suivantes</b>	
Demande de brevet		<input checked="" type="checkbox"/>	
Demande de certificat d'utilité		<input type="checkbox"/>	
Demande divisionnaire		<input type="checkbox"/>	
<i>Demande de brevet initiale</i> <i>ou demande de certificat d'utilité initiale</i>		N° _____ Date _____ N° _____ Date _____	
Transformation d'une demande de brevet européen <i>Demande de brevet initiale</i>		<input type="checkbox"/> N° _____ Date _____	
<b>TITRE DE L'INVENTION</b> (200 caractères ou espaces maximum) Dispositif et méthode perfectionnés de spatialisation du son.			
<b>4 DÉCLARATION DE PRIORITÉ</b> <b>OU REQUÊTE DU BÉNÉFICE DE</b> <b>LA DATE DE DÉPÔT D'UNE</b> <b>DEMANDE ANTÉRIEURE FRANÇAISE</b>		Pays ou organisation _____ N° _____ Date _____ Pays ou organisation _____ N° _____ Date _____ Pays ou organisation _____ N° _____ <input type="checkbox"/> S'il y a d'autres priorités, cochez la case et utilisez l'imprimé «Suite»	
<b>5 DEMANDEUR</b> (Cochez l'une des 2 cases)		<input checked="" type="checkbox"/> <b>Personne morale</b> <input type="checkbox"/> <b>Personne physique</b>	
Nom ou dénomination sociale		INRIA INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE	
Prénoms			
Forme juridique		Etablissement Public National à caractère Scientifique et Technologique	
N° SIREN		_____	
Code APE-NAF		_____	
Domicile ou siège	Rue	Domaine de Voluceau - Rocquencourt - BP 105	
	Code postal et ville	17 8 1 5 3 LE CHESNAIS Cedex	
	Pays	France	
Nationalité		française	
N° de téléphone (facultatif)		N° de télécopie (facultatif)	
Adresse électronique (facultatif)			
<input type="checkbox"/> S'il y a plus d'un demandeur, cochez la case et utilisez l'imprimé «Suite»			

Remplir impérativement la 2<sup>ème</sup> page



# BREVET D'INVENTION CERTIFICAT D'UTILITÉ

REQUÊTE EN DÉLIVRANCE  
page 2/2

**BR2**

DB 540 W / 210502

REMISE DES PIÈCES DATE <b>26 NOV 2003</b> LIEU <b>75 INPI PARIS 34 SP</b> N° D'ENREGISTREMENT <b>0313875</b> NATIONAL ATTRIBUÉ PAR L'INPI	
---	--

<b>6 MANDATAIRE (s'il y a lieu)</b>	
Nom	PLAÇAIS
Prénom	Jean-Yves
Cabinet ou Société	Cabinet NETTER
N° de pouvoir permanent et/ou de lien contractuel	
Adresse	Rue 36 avenue Hoche
	Code postal et ville <b>75 008 PARIS</b>
	Pays France
N° de téléphone (facultatif)	01 58 36 44 22
N° de télécopie (facultatif)	01 42 25 00 45
Adresse électronique (facultatif)	
<b>7 INVENTEUR (S)</b>	
Les inventeurs sont nécessairement des personnes physiques	
Les demandeurs et les inventeurs sont les mêmes personnes	<input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non : Dans ce cas remplir le formulaire de Désignation d'inventeur(s)
<b>8 RAPPORT DE RECHERCHE</b>	
Uniquement pour une demande de brevet (y compris division et transformation)	
Établissement immédiat ou établissement différé	<input checked="" type="checkbox"/> Établissement immédiat <input type="checkbox"/> Établissement différé
Paiement échelonné de la redevance (en deux versements)	Uniquement pour les personnes physiques effectuant elles-mêmes leur propre dépôt <input type="checkbox"/> Oui <input type="checkbox"/> Non
<b>9 RÉDUCTION DU TAUX DES REDEVANCES</b>	
Uniquement pour les personnes physiques	
<input type="checkbox"/> Requête pour la première fois pour cette invention (joindre un avis de non-imposition) <input type="checkbox"/> Obtenue antérieurement à ce dépôt pour cette invention (joindre une copie de la décision d'admission à l'assistance gratuite ou indiquer sa référence) : AG	
<b>10 SÉQUENCES DE NUCLEOTIDES ET/OU D'ACIDES AMINÉS</b>	
<input type="checkbox"/> Cochez la case si la description contient une liste de séquences	
Le support électronique de données est joint <input type="checkbox"/> La déclaration de conformité de la liste de séquences sur support papier avec le support électronique de données est jointe <input type="checkbox"/>	
Si vous avez utilisé l'imprimé «Suite», indiquez le nombre de pages jointes	
<b>11 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE (Nom et qualité du signataire)</b>	
N° Conseil 92-1197 (B) (M) Jean-Yves PLAÇAIS	
VISA DE LA PRÉFECTURE OU DE L'INPI 	

Dispositif et méthode perfectionnés de spatialisation du son

5 L'invention concerne le domaine du traitement du son.

La technique actuelle du traitement du son permet d'ajouter à la présentation d'une scène, en particulier en 3D sur écran, un son spatialisé de façon à améliorer de manière significative pour un spectateur le réalisme et le sens de l'immersion dans la scène. Cette technique est  
10 appropriée pour le traitement en temps réel d'un nombre limité de sources sonores dans la scène.

Les scènes, notamment virtuelles, deviennent de plus en plus complexes; en d'autres termes, le nombre de sources sonores dans une scène augmente. Ainsi, traiter en temps réel ces  
15 nombreuses sources sonores et effectuer un rendu sonore spatialisé pour ce grand nombre de sources sonores est souvent impossible en raison du coût élevé du traitement du signal.

L'invention vient améliorer la situation.

20 L'invention concerne un dispositif informatique comprenant une mémoire apte à stocker des signaux audio en partie pré-enregistrés, correspondant chacun à une source définie par des données de position spatiale,  
un module de traitement pour traiter ces signaux audio en temps réel en fonction des données de position spatiale.

25 Selon une caractéristique principale de l'invention, le module de traitement est apte à calculer des paramètres de niveau de puissance instantané à partir des signaux audio, les source correspondantes étant définies par lesdits paramètres de niveau de puissance instantané, le module de traitement comprend un module de sélection apte à regrouper  
30 certains des signaux audio en un nombre variable de groupes de signaux audio, et le module de traitement est apte à calculer des données de position spatiale représentatives d'un groupe

de signaux audio en fonction des données de position spatiale et des paramètres de niveaux de puissance instantanés de chaque source correspondante.

Le dispositif informatique selon l'invention peut comprendre de nombreuses caractéristiques

5 supplémentaires qui pourront être prises séparément et/ou en combinaison :

- le module de sélection apte, préalablement à la construction de groupes de signaux audio, à sélectionner les signaux audio inaudibles en fonction des paramètres de niveaux de puissance instantanés comprenant un niveau de puissance et un seuil de masquage pour chaque source et à conserver les seuls signaux audio audibles;
- 10 - les paramètres de niveau de puissance sont calculées pour chaque source à partir de la densité spectrale de puissance instantanée pré-calculé à partir des signaux audio en partie pré-enregistrés;
- le module de traitement est apte à traiter chaque groupe de signaux audio en un signal audio de pré-mixage et pour rassembler les signaux audio de pré-mixage afin d'obtenir un signal
- 15 de mixage audible par l'auditeur;
- le module de traitement comprend un processeur vidéo apte à transformer le groupe de signaux audio en un groupe de signaux vidéo texturés, à traiter chaque signal vidéo texturé du groupe selon des paramètres de modification du son, et à rassembler et transformer les signaux en un signal audio de pré-mixage;
- 20 - les paramètres de modification du son comprennent un paramètre d'atténuation du son et/ou un paramètre de délai de propagation du son;
- le module de sélection est apte à construire, à partir d'un premier groupe de signaux audio et de données calculées de position spatiale du groupe, deux groupe de signaux audio et à calculer les données de position spatiale d'un représentant de chacun de ces deux groupes,
- 25 - le module de sélection est apte à déterminer, à partir du premier groupe de signaux audio, de leurs sources correspondantes et de données calculées de position spatiale du représentant du premier groupe, une source pour laquelle la somme des distances d'erreur calculées entre la position spatiale de cette source et celles des autres sources du groupe est minimale et pour attribuer les signaux audio du premier groupe et leurs sources correspondantes à l'une des
- 30 positions spatiales, parmi les données calculées de position spatiale du représentant du



premier groupe et les données de position spatiale de la source déterminée, en fonction d'évaluations de distance d'erreur, de façon à former deux groupes;

- le module de sélection est apte à effectuer une évaluation de distance d'erreur pour un signal audio du premier groupe et sa source correspondante consistant à évaluer d'une part
- 5 la distance d'erreur entre les données de position spatiale de cette source et les données calculées de position spatiale du représentant du premier groupe et d'autre part la distance d'erreur entre les données de position spatiale de cette source et les données de position spatiale de la source déterminée puis à évaluer la distance d'erreur minimale entre les deux et le module de sélection étant apte à attribuer le signal audio et sa source correspondante
- 10 aux données de position spatiale de la source déterminée ou du représentant du premier groupe correspondant à la distance d'erreur minimale,
- les données de position spatiale de la source déterminée correspondent aux données de position spatiale du représentant d'un deuxième groupe;
- le module de sélection est apte à calculer les données de position spatiale de chaque
- 15 représentant de groupe en fonction de paramètres de niveau de puissance de chaque source attribuée au groupe
- le module de sélection est apte à recalculer les données de position spatiale du représentant de chacun des deux groupes en déterminant une source pour laquelle la somme des distances d'erreur entre la position spatiale de cette source et celles des autres sources du groupe est
- 20 minimale et le module de sélection est en outre apte à ré-attribuer les sources à l'un ou l'autre des représentants d'un des deux groupes en fonction de ladite évaluation de distance d'erreur minimale;
- le module de sélection est apte à recalculer les données de position spatiale du représentant de chacun des deux groupes et à ré-attribuer les sources à l'un ou l'autre des représentants
- 25 d'un des deux groupes jusqu'à ce que la somme des distances d'erreur entre les représentants des deux groupes et leurs sources atteigne un minimum;
- le module de sélection est apte à diviser un groupe jusqu'à ce qu'un nombre déterminé de groupes soit obtenu ou jusqu'à ce que la somme des distances d'erreur entre les représentants des groupes et leurs sources atteigne un seuil déterminé.

L'invention concerne également un procédé de traitement de signaux audio en partie pré-enregistrés correspondant chacun à une source, comprenant les étapes consistant à

- a. calculer des paramètres de niveau de puissance instantané à partir des signaux audio, les source correspondantes étant définies par ces paramètres et par des données de position spatiale,
- b. regrouper certains des signaux audio en un nombre variable de groupes de signaux audio et calculer des données de position spatiale représentatives de chaque groupe de signaux audio en fonction des données de position spatiale et des paramètres de niveau de puissance instantané de chaque source correspondante,
- c. traiter ces signaux audio par groupe en temps réel en fonction de données de position spatiale représentatives du groupe.

D'autres caractéristiques et avantages de l'invention apparaîtront à l'examen de la description détaillée ci-après, ainsi que des dessins annexés sur lesquels:

- la figure 1 représente un dispositif informatique dans lequel est réalisé l'invention,
- la figure 2 illustre des éléments matériels dans leur utilisation pour le traitement de signaux audio de l'art antérieur,
- la figure 3 illustre des éléments matériels dans leur utilisation pour le traitement de signaux audio de l'invention,
- la figure 4 est un ordinogramme illustrant le procédé de traitement de signaux audio de l'invention,
- la figure 4A est un ordinogramme détaillant une étape de division par groupe du procédé de la figure 4,

- la figure 4B est un ordinogramme détaillant une étape de traitement des signaux par groupe du procédé de la figure 4,

5 - la figure 5 représente schématiquement la comparaison entre l'utilisation de coordonnées cartésiennes et polaires pour la détermination du positionnement d'une source sonore fictive remplaçant deux sources sonores réelles,

- la figure 6 illustre le traitement d'un signal audio sous forme d'un signal vidéo par un processeur graphique 3D,

10

- la figure 7 illustre le traitement d'un signal en un signal comprimé temporellement et atténué,

15 - la figure 8 illustre, pour une configuration de quatre groupes de sources, deux échogrammes des signaux de pré-mixage de chaque groupe obtenus différemment.

L'annexe 1 présente les formules mathématiques utilisées pour la réalisation de l'invention.

L'annexe 2 présente les différentes variables utilisées et leur signification.

20 Les dessins et les annexes contiennent, pour l'essentiel, des éléments de caractère certain. Ils pourront donc non seulement servir à mieux faire comprendre la description, mais aussi contribuer à la définition de l'invention, le cas échéant.

25 La figure 1 représente un dispositif informatique comprenant une unité centrale 4 reliée à des périphériques tels qu'un écran 2, un clavier 5, une souris, un dispositif de hauts-parleurs 6 et autres. Ce dispositif informatique est utilisé pour la présentation visuelle dynamique sur écran d'un environnement (appelé aussi "scène") définissant différentes sources sonores et pour la restitution par haut-parleurs des sons provoqués par ces dernières. L'unité centrale comprend ainsi différents composants matériels aptes à traiter les signaux audio comme  
30 décrits en référence à la figure 2.

Ainsi, il est connu d'utiliser un processeur audio (ou module de traitement) relié à une mémoire 8 et à un dispositif de haut-parleurs 28. Le processeur audio 10 peut faire partie d'une carte son et est alors appelé DSP ("Digital Signal Processor" pour Processeur de Signal Numérique). Le processeur audio reçoit les signaux numériques provenant du processeur de la carte mère et les convertit en signaux analogiques transformés par haut-parleurs en sons. Les processeurs DSP plus performants permettent de traiter les signaux numériques en ajoutant des distorsions de signal, des échos (appelés réverbérations) par exemple. Certaines cartes mères intègrent elles-mêmes une carte son munie du processeur DSP. Ainsi, dans le cas de la figure 2, le processeur audio travaille avec des données de signaux audio 14 et avec des données de positions spatiales d'un utilisateur (appelé aussi "auditeur", ou "spectateur") par rapport à la scène et de sources sonores 16 enregistrés en mémoire 8. Les signaux audio sont chacun émis par une source sonore ayant une position spatiale définie dans une scène ou environnement présenté à l'écran. De façon connue, une position spatiale peut être représentée en mémoire par un triplet de coordonnées cartésiennes, polaires ou autres. La définition de la position spatiale d'un auditeur donné permet également d'obtenir un rendu audio pour ce dernier.

Comme indiqué sur la figure 2 et de façon connue, le processeur audio reçoit les données de la mémoire 8, c'est-à-dire chaque donnée de signal audio représenté par une flèche 14-i (i étant un entier positif représentant l'un des signaux audio) et les données de position des sources correspondantes et de l'auditeur. Les signaux audio sont traités par le processeur audio. Ce traitement se traduit par l'ajout d'effets 18 comprenant des opérations qui doivent être réalisées pour chaque signal audio d'entrée comme par exemple l'ajout de l'effet Doppler, l'ajout d'un délai, l'atténuation par la distance, l'ajout des effets d'occlusion/d'obstruction, de directivité. D'autres effets comme les effets de positionnement 22 de chaque signal de source dans la scène peuvent être ajoutés (sons provenant d'une source lointaine ou d'une source proche de l'auditeur, rendu de la direction de provenance des sons aux oreilles de l'auditeur). Les signaux audio sont ensuite soumis à un mixage 24 correspondant à la sommation des signaux ainsi traités. Après l'ajout des effets 18, les signaux peuvent être sommés en un signal soumis à certains effets, par exemple un effet de

réverbération. Le signal résultant est ajouté à la sommation des signaux spatialisés grâce au module de mixage 24 afin d'obtenir un signal sonore final. Le processeur audio traite les signaux audio en temps réel en fonction d'une donnée de position spatiale d'un auditeur.

5 Ainsi, le processeur audio 10 délivre un signal analogique transformé en son et diffusé par le dispositif haut-parleur 28. Ce dispositif informatique permet d'obtenir un rendu sonore spatialisé améliorant le sens de réalisme et d'immersion dans la scène ou l'environnement présenté à l'écran. Des exemples de cartes sonores connues sont détaillés sur les pages internet suivantes :

- 10 [1] Creative Labs Soudblaster©. <http://www.soundblaster.com>  
 [2] Direct X homepage, ©microsoft  
 [3] Environmental audio extensions: EAX 2.0 Creative ©

15 Toutefois, la technique décrite ci-dessus parvient à ses limites lorsqu'un grand nombre de sources sonores est défini dans la scène. Le traitement de ce grand nombre de sources sonores devient impossible en raison du coût du traitement des nombreux signaux.

Il est intéressant de noter que le dispositif informatique décrit précédemment est en général limité à des sources sonores ponctuelles. Hors pour obtenir un rendu sonore réaliste de sources sonores étendues (c'est-à-dire non ponctuelles, un train par exemple) il est possible d'échantillonner la surface ou le volume définissant la source en une collection de source ponctuelles. Un désavantage d'une telle approche est qu'elle multiplie rapidement le nombre de sources à traiter. On rencontre un problème similaire lorsque les réflexions ou diffractions du son sur les parois de l'environnement virtuel doivent être modélisées sous forme de "source-images". Ceci est présenté dans les articles :

20 25

- [4] J.B. Allen et D.A Berkley, Image method for efficiently simulating small room acoustics, *Journal of the Acoustical Society of America*, n°4, Vol. 65, 1979.  
 [5] J.Borish, Extension of the image model to arbitrary polyhedra, *Journal of the Acoustical Society of America*, n°6, vol.75, 1984.

[6]N.Tsingos et T.Funkhouser et A.Ngan et I. Carlbom, Modeling Acoustics in Virtual Environments using the Uniform Theory of Diffraction, *ACM Computer Graphics, SIGGRAPH'01 Proceedings*, p.545-552, août 2001.

- 5 Une solution technique a été présentée par Herder se basant sur le regroupement de sources sonores et sur la sélection d'un nombre fixe de sources sonores représentatives de l'ensemble des sources sonores. Toutefois, la technique de Herder reste coûteuse et peu adaptable à un budget donné. Cette technique est décrite dans l'ouvrage suivant :
- [7] Jens Herder. Optimization of sound spatialization resource management through clustering. *The Journal of Three Dimensional Images, 3D-Forum Society*, 13(3):59-70, 10 Septembre 1999.

Ainsi, l'invention présente une solution technique permettant le rendu audio-visuel d'environnements contenant des centaines de sources sonores à un coût moindre que les 15 solutions antérieures. Avantageusement, l'invention permet un rendu audio-visuel avec réverbération et effets dépendant de la fréquence, sur des systèmes grand-public.

Un exemple de réalisation du dispositif de l'invention est détaillé sur la figure 3. Le dispositif comprend une mémoire 108 permettant de stocker des données de signaux audio 20 114 et de positions des sources sonores correspondantes ainsi que de position de l'auditeur 116. Cette mémoire travaille en relation avec un module de traitement 110 comprenant un module de sélection 120, un processeur vidéo 130 et un processeur audio 140. A titre d'exemple, le dispositif mettant en oeuvre le procédé de l'invention peut être un ordinateur *PC Xeon 1.8 Ghz* comprenant une carte son qui peut être une carte *Soundblaster Audigy* ou 25 une carte *SoundMax* et une carte vidéo qui peut être une carte *GeForce 4600Ti* ou une carte *ATI Radeon Mobility 5700*.

Le traitement des signaux audio en fonction des positions des sources sonores et de la position de l'auditeur est décrit sur l'ordinogramme de la figure 4 en correspondance avec 30 la figure 3.

Préalablement au traitement des signaux par le module de traitement, des informations de type densité spectrale de puissance instantanée PSD et seuil de puissance de masquage  $M$  128 sont calculées par le module de traitement pour chaque position de source sonore stockée en mémoire. Plusieurs expressions pour le calcul du seuil de puissance de masquage

5 sont connues des applications de codage audio perceptuel (PAC), par exemple dans le standard de la couche III (mp3) de MPEG-1. Ces expressions peuvent être trouvées dans les ouvrages suivants

[8] K. Brandenburg. mp3 and aac explained. *AES 17th International Conference on High-Quality Audio Coding*, Septembre 1999,

10 [9] R. Rangachar. Analysis and improvement of the MPEG-1 audio layer III algorithm at low bit-rates. *Master thesis, Arizona State University*, Decembre 2001.

Ce seuil de puissance de masquage calculé correspond à la puissance maximale d'un bruit pouvant être masqué par le signal. Dans le cas de l'invention, les seuils de puissance de masquage peuvent être dynamiquement calculés pour un grand nombre de sources.

15 Toutefois, comme les données de signaux audio sont pré-enregistrés et non pas dynamiquement évalués, les seuils de puissance de masquage  $M$  sont dynamiquement calculés à partir d'information de tonalité  $T$  (ou index de tonalité) pouvant être pré-calculés et pré-enregistrés en 114, puis dynamiquement accessibles. Il peut en aller de même des densités spectrales de puissance instantanée PSD qui permettent de pré-calculer les informations de tonalité  $T$ .

20

Il est également envisageable d'évaluer les densités spectrales de puissance instantanée PSD à la volée si l'intégralité du signal n'est pas disponible à l'avance (par exemple quand les données audio sont synthétisées ou transmises sous forme de flux dit "streaming") de façon à calculer les informations de tonalité  $T$ . On parle alors de signaux en partie pré-enregistrés.

25 Néanmoins cette évaluation à la volée nécessite une plus forte puissance de calcul.

A titre d'exemple uniquement, ces informations sont calculées pour chaque signal audio et plus précisément pour trois composantes pré-calculées de chaque signal audio correspondant à trois bandes de fréquence du spectre audio audible. Le nombre de trois bandes de

30 fréquence n'est en aucun cas limitatif et pourrait par exemple être de vingt-cinq bandes. Ces

bandes de fréquence audibles sont, dans cet exemple,  $f_1 = [0-500 \text{ Hz}]$  basses fréquences,  $f_2 = [500-2000 \text{ Hz}]$  moyennes fréquences et  $f_3 = [+ \text{ de } 2000 \text{ Hz}]$  hautes fréquences. Les seuils de puissance de masquage  $M$  et les densités spectrales de puissance instantanée  $PSD$  sont calculés à partir des techniques décrites dans les ouvrages suivants

- 5 [10] K.Brandenburg. mp3 and aac explained., *AES 17th International Conference on High-Quality Audio Coding*, Septembre 1999.

[11] E.M.Painter and A.S.Spanias. A review of algorithms for perceptual coding of digital audio signals. *DSP-97*, 1997

- 10 [12] R.Rangachar. Analysis and improvement of the MPEG-1 audio layer III algorithm at low bit-rates. *Master thesis, Arizona State University*, December 2001.

[13] Ken Steiglitz. *A DSP Primer with applications to digital audio and computer music*. Addison Wesley, 1996.

- 15 Pour chaque bande de fréquence, une transformée de Fourier est calculée à partir des techniques des ouvrages [8], [9] et

[14] E.M. Painter and A.S.Spanias. A review of algorithms for perceptual coding of digital audio signals. *DSP-97*, 1997.

- Pour chaque bande de fréquence  $f$ , la distribution du spectre de puissance instantanée  $PSD_t(f)$  est calculée pour chaque image  $t$ . Pour le calcul du seuil de masquage  $M$ , on se reportera aux équations A1 à A3 de l'annexe A.
- 20

- A un instant donné, le module de sélection reçoit les signaux audio 114, les informations 128 de type seuil de masquage ainsi que le spectre de puissance instantanée  $PSD$ . Avec ces données, le module de sélection effectue un tri entre les signaux et isole les sources inaudibles à l'étape 200 de la figure 4.
- 25

- Pour ce faire, le module de sélection estime à l'instant  $T$  le volume perceptif  $L_k^T$  du signal audio, comme indiqué par l'équation A4, de chaque source sonore  $k$  et pour l'ensemble des bandes de fréquences  $f$ . Comme indiqué à l'équation A5, ce volume perceptif est fonction
- 30



du niveau de puissance de chaque bande de fréquence  $f$  à un instant  $T-\delta$ , instant tenant compte du délai de propagation du signal entre la position de la source et la position de l'auditeur, et de la contribution  $\alpha(f)$  différente au volume perceptif de chaque niveau de puissance  $P(f)$ . Le niveau de puissance de chaque bande de fréquence  $f$  est calculé à partir

5 de la distribution spectrale de puissance instantanée PSD de la source à l'instant  $T-\delta$ , de l'atténuation  $A$  dépendant de la distance, de l'occlusion, du modèle de directivité de la source par exemple. Ce volume perceptif instantané peut être moyenné sur les instants précédents (par exemple les dix instants  $T$  précédents). On parle de "paramètres de niveau de puissance" pour englober le seuil de puissance de masquage et les paramètres dépendants

10 des niveaux de puissance, c'est-à-dire les niveaux de puissance eux-mêmes et les volumes perceptifs par exemple. Une source est définie par sa position spatiale et ses paramètres de niveau de puissance calculés par le module de traitement 110 de la figure 3. A chaque instant  $T$ , le module de sélection 120 trie les sources sonores dans l'ordre décroissant des résultats obtenus par le calcul du critère de l'équation A6 combinant le volume perceptif et le seuil

15 de masquage. Le critère A6 peut donc être considéré comme une quantification de l'importance perceptive de chaque source dans la scène sonore globale.

Après avoir calculé le niveau de puissance global de la scène  $P_o$  pour l'ensemble des sources en A7 à un instant donné, l'algorithme A8 est effectué à cet instant donné et pour chaque

20 source  $S_k$  de manière à sélectionner et éliminer les sources inaudibles. L'algorithme A8 insère progressivement les sources  $S_k$  par ordre décroissant d'importance dans le mixage courant  $P_{mix}$ . Le niveau de puissance  $P_k$  de la source est retiré de la puissance globale de la scène  $P_o$  et est ajouté à la puissance courante du mixage  $P_{mix}$ , le seuil de puissance de masquage  $M_k$  de la source est ajouté au seuil de puissance de masquage courant  $T_{mix}$  du

25 mixage. L'algorithme A8 est itéré pour chaque source  $S_k$  tant que les deux conditions suivantes sont vérifiées :

1. la puissance globale courante de la scène est supérieure à la différence entre la puissance courante du mixage et le seuil de puissance de masquage courant du mixage,
2. la puissance globale restante est supérieure au seuil absolu d'écoute (ATH).

Sur la figure 3, les signaux sont représentés par des flèches reçus dans le module de sélection et les signaux inaudibles sont représentés par des flèches qui s'arrêtent dans le module de sélection 120 en une croix. Ces opérations se répètent successivement pour chaque instant.

- 5 Avec les sources audibles, le module de sélection détermine le nombre N de groupes de signaux audio audibles (ou de sources audibles) possible à réaliser à l'étape 202.

Le nombre N de groupes peut être directement pré-déterminé par l'utilisateur, enregistré et lu par le module de sélection par exemple, ou résulter de la valeur du seuil d'erreur définie ultérieurement en A10, valeur fixée par l'utilisateur. Un groupe de source peut être spatialisé en utilisant un canal audio de la carte son (DSP). Le nombre N de groupes pourra donc être choisi comme étant égal au nombre maximum de canaux pouvant être spatialisé par la carte son. Si la spatialisation, c'est-à-dire le traitement positionnel du son, doit être effectué à l'aide du processeur central, une évaluation du coût du calcul d'un groupe peut permettre à l'utilisateur de déterminer quel nombre N de groupes construire. Il est possible également de maintenir dynamiquement une évaluation du coût des calculs effectués pour chaque groupe (par exemple en évaluant le temps processeur nécessaire) et d'adapter le nombre N de groupes en fonction. Ce nombre N est donc variable selon les valeurs rentrées par l'utilisateur, selon le nombre de canaux de la carte son ou selon l'évaluation de coûts par exemple.

A l'étape 204, le module de sélection est apte à regrouper en N groupes les signaux audio. Le module de traitement est apte à calculer une position spatiale représentative pour chaque groupe de signaux audio en fonction de la position spatiale et du volume perceptif de chaque source correspondante.

Le procédé de l'étape 204 sera plus particulièrement détaillé en référence à la figure 4A par la suite. Le procédé de l'étape 204 peut également effectuer un regroupement de signaux audio en utilisant un autre procédé que celui détaillé en référence à la figure 4A. Ainsi, il est

possible de déterminer les N représentants choisis parmi les sources en utilisant une heuristique telle que présentée dans l'ouvrage

- [14] Hochbaum, D. and Shmoys, D. A best possible heuristic for the k-center problem.  
5 Mathematics of Operations Research, 1985.

Les N groupes sont ainsi formés en affectant chaque source au représentant le plus proche au sens de la métrique définie dans l'équation A9 détaillée plus loin.

- 10 A l'étape 206, les signaux audio de chaque groupe sont traités afin d'obtenir un signal audio de pré-mixage par groupe. L'obtention d'un signal de pré-mixage par groupe sera explicitée en relation avec la figure 4B détaillant l'étape 206. De manière avantageuse et en référence à la figure 3, l'étape de pré-mixage des signaux par groupe se déroule dans le processeur vidéo 130 dans un module de pré-mixage 132. Par pré-mixage, on entend tout d'abord les  
15 opérations qui doivent être réalisées pour chaque signal audio d'entrée comme par exemple l'ajout de l'effet doppler, l'ajout d'un délai, l'atténuation par la distance, des effets d'occlusion/d'obstruction, de directivité, ainsi que la somme des signaux ainsi traités dans chaque groupe. Le pré-mixage peut également comprendre la sommation de tous les signaux de tous les groupes afin d'ajouter un effet de réverbération 146 sur ce signal de sommation  
20 ( $\Sigma$ ). Ensuite, le processeur audio 140 reçoit un signal audio de pré-mixage pour chaque groupe, et le signal de sommation ( $\Sigma$ ). Le processeur audio peut ajouter des effets de réverbération 146 sur le signal de sommation. Le processeur audio applique un effet de positionnement 142 sur chaque signal audio de pré-mixage avant de mixer ces derniers entre-eux ainsi que le signal résultant du module de réverbération 146 afin d'obtenir un signal  
25 audio de mixage audible par l'auditeur à l'étape 208.

Par mixage, on entend, après les opérations de positionnement des signaux dans la scène, la sommation finale des opérations de positionnement et des effets de réverbération si il y a lieu.

L'étape 204 est maintenant détaillée en référence à la figure 4A.

Dans un premier temps, le regroupement des sources en groupes s'effectue en construisant un premier groupe unique réunissant les sources audibles, puis ce groupe est successivement  
5 découpé pour obtenir le nombre de groupes voulu. Dans le cas où le nombre de groupes est plus grand que le nombre de sources disponibles, chaque source représentera un groupe.

A l'étape 2000, le module de sélection définit un premier groupe unique réunissant les sources audibles et calcule la position spatiale du représentant C1 du groupe. Cette position  
10 spatiale correspond à l'évaluation du centroïde à partir du jeu de positions spatiales des sources émettant les signaux audio. Dans l'exemple de l'invention et comme illustré par la figure 5, il est intéressant d'utiliser les coordonnées polaires pour définir les positions spatiales des sources S1 et S2 éloignées de l'auditeur afin de déterminer un centroïde polaire CP du représentant du groupe et non un centroïde cartésien CC. En effet, le centroïde  
15 cartésien CC du représentant du groupe est très proche de l'auditeur AU et ne permet pas de conserver la distance entre les sources (S1 et S2) et l'auditeur. Au contraire, le centroïde polaire CP du représentant du groupe conserve la distance avec l'auditeur AU et donc le délai de propagation du signal jusqu'à l'auditeur. De façon à déterminer la position spatiale du représentant C1 du groupe à la manière d'un barycentre, le volume perceptif de chaque  
20 source peut être associé à ses coordonnées spatiales comme indiqué en A11.

A l'étape 2002, une source Si du groupe est choisie telle que ses données minimisent une fonction d'erreur globale définie en A10. En effet, un représentant de groupe doit assurer que les distorsions acoustiques sont minimales lorsqu'il est utilisé pour spatialiser le signal. La  
25 fonction d'erreur globale est la somme des distances d'erreurs ou "métriques d'erreurs" pour toutes les sources du groupe. Ces distances d'erreurs ou "métriques d'erreurs" sont définies en A9 comme la somme de deux termes de déviation spatiale entre une source et le représentant du groupe. Ainsi, l'étape 2002 consiste à déterminer, à partir du premier groupe de signaux audio, de leurs sources correspondantes et de données calculées de position  
30 spatiale du représentant C1 du premier groupe, une source pour laquelle la somme des

distances d'erreur calculées entre la position spatiale de cette source et celles des autres sources du premier groupe est minimale.  $C$  et  $S_k$  utilisées dans  $A_9$  correspondent respectivement à un premier et un deuxième vecteur, dans un repère centré sur la position courante de l'auditeur, ayant pour coordonnées spatiales cartésiennes respectivement celles du centroïde  $C$  et celles de la source  $S_k$ . Les deux termes de la somme comprennent un terme de déviation de distance et un terme de déviation d'angle. La contribution du volume perceptif de la source permet d'assurer une distance d'erreur minimale pour les sources ayant un fort volume perceptif. A titre d'exemple uniquement, les paramètres  $\gamma$  et  $\beta$  peuvent prendre les valeurs 1 et 2 respectivement pour équilibrer l'importance des termes de déviation entre-eux.

La source  $S_i$  choisie devient le nouveau représentant  $C_2$  d'un deuxième groupe à constituer. A l'étape 2004, les signaux audio du groupe et les sources correspondantes sont attribuées soit au représentant  $C_1$ , soit au représentant  $C_2$  selon un critère donné. Ainsi, l'étape 2004 consiste à attribuer les signaux audio du premier groupe et leurs sources correspondantes à l'une des positions spatiales, parmi les données calculées de position spatiale du représentant  $C_1$  du premier groupe et les données de position spatiale de la source  $S_i$  déterminée, en fonction d'évaluations de distance d'erreur, de façon à former les deux groupes. La distance d'erreur entre la position spatiale d'une source  $S_k$  du groupe et la position spatiale du représentant  $C_1$  du groupe est comparée à la distance d'erreur entre la position spatiale de la même source et la position spatiale du représentant  $C_2$  (correspondant à la source  $S_i$ ). La distance d'erreur minimale permet de déterminer le représentant auquel le signal audio et la source correspondante seront attribués. Plus précisément, le signal audio et sa source correspondante sont attribués aux données de position spatiale de la source  $S_i$  déterminée (correspondant au représentant  $C_2$ ) ou du représentant  $C_1$  du premier groupe correspondant à la distance d'erreur minimale (2004).

Une fois l'attribution des signaux audio et leurs sources aux représentants  $C_1$  ou  $C_2$  effectuée, la position spatiale des représentants  $C_1$  et  $C_2$  est recalculée selon  $A_{11}$  pour optimisation à l'étape 2006. A l'étape 2008, les représentants  $C_1$  et  $C_2$  ayant de nouvelles

positions spatiales, une nouvelle attribution des signaux audio et leurs sources aux représentants C1 et C2 est effectuée selon le même critère de distance d'erreur minimale qu'à l'étape 2002. Les étapes 2006, c'est-à-dire recalculer les données de position spatiale du représentant de chacun des deux groupes, et 2008, c'est-à-dire ré-attribuer les sources à

5 l'un ou l'autre des représentants d'un des deux groupes, sont répétées jusqu'à ce qu'un critère soit vérifié à l'étape 2010. Dans la réalisation présentée, le critère de l'étape 2010 est que la somme des erreurs globales pour les représentants des deux groupes atteigne un minimum local de la fonction d'erreur A10. En d'autres termes, ce critère de l'étape 2010 est que la somme des distances d'erreur entre les représentants des deux groupes et leurs

10 sources atteigne un minimum.

Après l'obtention de groupes dont les représentants ont des positions spatiales optimisées vis à vis des sources de chaque groupe, il est possible de re-diviser un des groupes en deux groupes de façon itérative (retour à l'étape 2002). Le groupe à diviser peut être choisi parmi

15 tous les groupes courants, par exemple celui dont l'erreur A10 est la plus importante. La subdivision est effectuée jusqu'à l'obtention du nombre de groupes souhaité ou jusqu'à ce que l'erreur globale, c'est-à-dire la somme des erreurs A10 pour chaque groupe, soit inférieure à un seuil prédéfini par l'utilisateur.

20 La figure 4B reprend en détail l'étape 206 de la figure 4. Les signaux audio sont reçus en groupes par le processeur vidéo. Comme vu précédemment et illustré sur la figure 6, chaque signal audio SO1 a été décomposé en trois composantes R, G, B pré-calculées correspondant à trois bandes de fréquence du spectre audio audible. Toutefois, d'autres bandes de fréquence que celles déjà utilisées peuvent être utilisées à l'étape 206. A l'étape 2020, dans

25 le processeur vidéo, ces composantes R, G, B sont chargées en mémoire sous forme d'une collection de tronçons texturés 1D. Ainsi, le signal vidéo SV1 résulte du filtrage du signal audio SO1 sous forme de deux lignes texturées, l'une pour la partie positive du signal, l'autre pour la partie négative du signal, chaque ligne comprenant une collection de tronçons texturés. Les textures possibles des tronçons peuvent correspondre, de façon non limitative,

30 à une variation de contrastes monochromatiques ou à une variation du noir au blanc, comme

illustrée. Selon la figure 6, pour la ligne positive du signal vidéo, plus le signal audio prend une valeur élevée plus le tronçon correspondant a une texture claire et pour toutes les valeurs négatives du signal audio, les tronçons correspondants prennent une même texture foncée. Pour la ligne négative du signal vidéo, plus le signal audio prend une valeur négative dont la valeur absolue est élevée plus le tronçon correspondant a une texture claire et pour toutes les valeurs positives du signal audio, les tronçons correspondants prennent une texture foncée, en général une texture noire.

La représentation sous la forme de deux lignes texturées n'est pas limitative et peut se réduire à une ligne s'il est utilisé une mémoire vidéo acceptant les valeurs négatives du signal.

A l'étape 2022, le signal vidéo de chaque source est ensuite ré-échantillonné pour tenir compte de la variable du délai de propagation prenant une valeur différente selon l'emplacement de la source par rapport à l'auditeur. A l'étape 2024, le signal vidéo de chaque source est également atténué selon la distance entre la source et l'auditeur. Ces étapes 2022 et 2024 de modification du signal selon des paramètres de modification du son peuvent être réalisées en même temps ou dans un ordre différent de celui de la figure 4B. D'autres paramètres de modification du son pourraient être envisagés, par exemple l'atténuation pourrait être une fonction de la fréquence. La figure 7 illustre le ré-échantillonnage et l'atténuation du signal d'une source. Le signal audio SO2 (fonction du temps) est d'abord filtré pour obtenir un signal vidéo SV2 par exemple sous forme de deux lignes texturées (l'une pour la partie positive du signal audio, l'autre pour la partie négative du signal audio), le signal formant un premier ensemble de blocs texturés TBk et un deuxième ensemble de blocs texturés TBk+1. Le ré-échantillonnage des deux ensembles est effectué pour réduire le temps de propagation du signal en fonction du délai de propagation. Le signal peut être également atténué selon une atténuation dépendant de la bande de fréquence et/ou selon une atténuation dépendant de la distance source-auditeur, ou plus précisément, une atténuation dépendant de la distance source-auditeur corrigée par la distance source-représentant du groupe. A titre de comparaison, le signal audio SO2 et vidéo correspondant SV2 sont

présentés après ré-échantillonnage temporel et atténuation de l'amplitude sur la figure 7. Le signal audio SO2 est ainsi compressé temporellement et l'amplitude du signal est atténuée progressivement en fonction du temps. Les opérations 2022 et 2024 effectuées sur le signal vidéo SV2 (correspondant au signal audio SO2) permettent d'obtenir un signal vidéo SV3 (correspondant au signal audio SO3) compressé temporellement et atténué progressivement en fonction du temps. La compression temporelle du signal vidéo se traduit par exemple par une largeur diminuée des tronçons texturés afin d'obtenir deux ensembles de bloc LS1 et LS2. L'atténuation progressive en fonction du temps se traduit par exemple par une modulation des textures des tronçons.

10

A l'étape 2026, chaque signal vidéo est converti en un signal audio en effectuant tout d'abord une recombinaison des deux lignes du signal vidéo (parties positive et négative du signal). Pour chaque groupe, les signaux audio sont alors rassemblés en un signal audio unique relié au groupe de sources. Le signal audio obtenu par groupe est appelé signal audio de pré-mixage. La figure 8 illustre, pour un ensemble de groupes G1, G2, G3 et G4 de sources et un auditeur L, deux échogrammes H1 et H2 donnant la quantité d'énergie délivrée par groupe en fonction du temps délivrée à l'auditeur L. Le premier échogramme H1 illustre le cas du procédé de la figure 4B. Ainsi, chaque signal de chaque groupe fait l'objet individuellement des opérations 2022 et 2024 avant le rassemblement des signaux par groupe à l'étape 2026. Cet ordre des étapes permet d'obtenir une distribution d'énergie dans le temps pour chaque groupe tout en prenant en compte le délai de propagation et l'atténuation de chaque signal du groupe. L'échogramme H2 illustre le cas où les opérations 2022 et 2024 ont été effectuées après le rassemblement des signaux audio par groupe de sources, c'est-à-dire sur chaque signal représentant un groupe. Cet ordre des étapes permet d'obtenir une distribution d'énergie dans le temps pour chaque groupe mais cette fois en prenant en compte le délai de propagation et l'atténuation du signal représentatif des signaux du groupe. L'ordre des étapes peut être choisi selon le degrés de fine perception des sons voulu par l'auditeur. Il est clair que la mémoire utilisée et les temps de calcul seront moindre dans le cas de l'histogramme H2 mais que la perception des sons de l'auditeur sera moins fine que dans le cas de l'histogramme H1.

30



Ce processus peut être implémenté sur toute carte graphique accélérant les routines des bibliothèques graphiques standard "OpenGL" ou "Direct 3D". Les capacités des nouvelles cartes graphiques permettent à présent de travailler avec des micro-programmes exécutés chaque fois qu'un pixel est affiché ("pixel shader" ou "fragment programs"). Dans ce cas, il est possible de travailler avec des données signées et il n'est pas nécessaire de séparer les parties positives et négatives du signal. De plus dans ce cas, les opérations peuvent être effectuées avec une résolution étendue (32 bit flottant contre 8 bits entiers sur des cartes plus anciennes). Pour cela, il est possible d'utiliser le même algorithme que précédemment pour construire une texture dont chaque ligne correspond au signal SV2 de chaque source. Puis les lignes voulues sont additionnées pour chacun des groupes dans un micro-programme "pixel shader" en traçant une nouvelle ligne par groupe. L'accès aux lignes voulues et leur addition sont effectués dans le programme du "pixel shader".

Chaque signal audio de pré-mixage est relié au représentant d'un groupe qui représente une source fictive. Ces signaux audio de pré-mixage peuvent être utilisés par un système audio spatialisé standard pour rendre audible les sources de la scène visualisée. A titre d'exemple, la spatialisation peut être effectuée par un logiciel ou par une interface de programmation standard pour le rendu audio de jeux tel que *Direct Sound*. Dans ce dernier cas, une mémoire tampon 3D audio peut être créée pour stocker le signal de pré-mixage de chaque groupe. Chaque signal de pré-mixage est alors positionné aux coordonnées du représentant de son groupe, par exemple en utilisant la commande *SetPosition* de l'interface de programmation *Direct sound*. D'autres traitements comme celui de la réverbération artificielle peuvent également être utilisés lorsque proposés par le système audio spatialisé standard utilisé.

L'approche décrite introduit trois étapes principales utilisant une élimination perceptive des sources sonores inaudibles, un regroupement permettant de rendre un grand nombre de sources sur un nombre limité de canaux audio cablés et le matériel graphique pour effectuer les opérations de prémixage nécessaires.

Avec peu d'impact sur les performances graphiques, le procédé et le dispositif associé permettent d'exploiter les ressources matérielles des cartes sonores existantes tout en introduisant des possibilités de contrôle et traitement supplémentaires.

- 5 La mise en oeuvre du procédé décrit par un dispositif approprié permet d'obtenir un rendu audio-visuel de qualité pour un environnement virtuel complexe comprenant des centaines de sources mobiles, de personnages et d'objets animés.

- 10 L'invention pourrait également s'appliquer à un dispositif informatique comprenant une carte mère comprenant elle-même un processeur vidéo ou une carte vidéo et un processeur audio ou une carte son.

Annexe 1

- 5
- A1 
$$\text{SFM}_t(f) = 10 \log_{10} \left( \frac{\mu_g(\text{PSD}_t(f))}{\mu_d(\text{PSD}_t(f))} \right),$$
- A2 
$$T_t(f) = \frac{\max(\text{SFM}_t(f), 0)}{-60} - 1.$$
- A3 
$$M_t(f) = 31 * T_t(f) + 12 * (1 - T_t(f)),$$
- 10 A4 
$$L_k^T = \sum_f \alpha(f) P_k^{T-\delta}(f),$$
- A5 
$$P_k^{T-\delta}(f) = \text{PSD}_k^{T-\delta}(f) \times A_k^T(f)/r^2,$$
- 15 A6 
$$L_k^T \parallel 60 - M_k^{T-\delta} \parallel$$
- A7 
$$P_{\text{TOT}} = \sum_k P_k^{T-\delta}(f).$$
- 20 A8
- ```

Po = PTOT
while Po > Pmix - Tmix
  and Po > ATH do
    add source Sk
    Po = Pk
    Pmix = Pk
    Tmix = Mk
  end

```
- 25 A9 
$$d(C, S_k) = L_k^T \left( \beta \log_{10}(\|C\|/\|S_k\|) + \gamma \frac{1}{2} (1 + C \cdot S_k) \right),$$
- A10 
$$E_n = \sum_j d(C, S_j)$$
- 30 A11
- $$\rho_C = \sum_j L_j^T r_j / (\sum_k L_k^T),$$
- $$\theta_C = \theta(\sum_j L_j^T S_j / (\sum_k L_k^T)),$$
- $$\phi_C = \phi(\sum_j L_j^T S_j / (\sum_k L_k^T)).$$

Annexe 2

$C$  : représentant d'un groupe - par extension dans les formules mathématiques, vecteur de coordonnées spatiales du représentant d'un groupe

- 5  $S_k$  : source sonore dans une scène virtuelle - par extension dans les formules mathématiques, vecteur de coordonnées spatiales de la source sonore

$L_k^T$  : volume perceptif d'un signal audio d'une source sonore  $S_k$  à un instant  $T$

$\alpha(f)$  : poids contrôlant l'importance perceptive relative d'une bande de fréquence  $f$  donnée

$f$  : bande de fréquence d'un signal audio

- 10  $P_k^{T-\delta}(f)$  : estimation du niveau de puissance de chaque bande de fréquence  $f$  du signal audio d'une source sonore  $K$  à un instant  $T-\delta$

$\delta$  : délai de propagation du signal audio

$r$  : distance source sonore - auditeur

$c$  : vitesse du son

- 15  $A_k^T(f)$  : atténuation dépendante de la fréquence et résultant notamment de la distance, la direction de la source

$PSD_k^{T-\delta}$  : distribution du spectre de puissance instantanée

ATH : seuil absolu d'écoute

$P_{tot}$  : niveau de puissance totale de la scène

- 20  $SFMT(f)$  : mesure d'aplatissement spectral

$\mu_g$  : moyenne géométrique de la PSD sur toutes les fréquences

$\mu_a$  : moyenne arithmétique de la PSD sur toutes les fréquences

$T_t(f)$  : index de tonalité, niveau de bruit d'un signal

$M_t(f)$  : seuil de masquage (en dB)

- 25  $P_{mix}$  : puissance courante du mixage

Revendications

1. Dispositif informatique comprenant  
une mémoire (8, 108) apte à stocker des signaux audio (14, 114) en partie pré-enregistrés,  
5 correspondant chacun à une source définie par des données de position spatiale (16,116),  
un module de traitement (10,110) pour traiter ces signaux audio en temps réel en fonction  
des données de position spatiale,  
caractérisé en ce que  
le module de traitement (110) est apte à calculer des paramètres de niveau de puissance  
10 instantané à partir des signaux audio (114), les source correspondantes étant définies par  
lesdits paramètres de niveau de puissance instantané,  
en ce que le module de traitement (110) comprend un module de sélection (120) apte à  
regrouper certains des signaux audio en un nombre variable de groupes de signaux audio,  
et en ce que le module de traitement (110) est apte à calculer des données de position spatiale  
15 représentatives d'un groupe de signaux audio en fonction des données de position spatiale  
(116) et des paramètres de niveaux de puissance instantanés de chaque source correspon-  
dante.
2. Dispositif informatique selon la revendication 1, caractérisé en ce que le module de  
20 sélection (120) est apte, préalablement à la construction de groupes de signaux audio, à  
sélectionner les signaux audio inaudibles en fonction des paramètres de niveaux de puissance  
instantanés comprenant un niveau de puissance ( $P_k^{T-\delta}(f)$ ) et un seuil de masquage ( $Mt(f)$ )  
pour chaque source et à conserver les seuls signaux audio audibles.
- 25 3. Dispositif informatique selon la revendication 2, caractérisé en ce que les paramètres de  
niveau de puissance sont calculées pour chaque source à partir de la densité spectrale de  
puissance instantanée (PSD) pré-calculé à partir des signaux audio en partie pré-enregistrés.
4. Dispositif informatique selon l'une des revendications 1 à 3, caractérisé en ce que le  
30 module de traitement (110) est apte à traiter chaque groupe de signaux audio en un signal

audio de pré-mixage et pour rassembler les signaux audio de pré-mixage afin d'obtenir un signal de mixage audible par l'auditeur.

5. Dispositif informatique selon l'une des revendications précédentes, caractérisé en ce que
- 5 le module de traitement (110) comprend un processeur vidéo (130) apte à transformer le groupe de signaux audio en un groupe de signaux vidéo texturés, à traiter chaque signal vidéo texturé du groupe selon des paramètres de modification du son, et à rassembler et transformer les signaux du groupe en un signal audio de pré-mixage.
- 10 6. Dispositif informatique selon la revendication 5, caractérisé en ce que les paramètres de modification du son comprennent un paramètre d'atténuation du son et/ou un paramètre de délai de propagation du son.
7. Dispositif informatique selon l'une des revendications précédentes, caractérisé en ce que
- 15 le module de sélection (120) est apte à construire, à partir d'un premier groupe de signaux audio et de données calculées de position spatiale du groupe, deux groupe de signaux audio et à calculer les données de position spatiale d'un représentant de chacun de ces deux groupes.
- 20 8. Dispositif informatique selon la revendication 7, caractérisé en ce que le module de sélection (120) est apte à déterminer, à partir du premier groupe de signaux audio, de leurs sources correspondantes et de données calculées de position spatiale du représentant du premier groupe, une source pour laquelle la somme des distances d'erreur calculées entre la position spatiale de cette source et celles des autres sources du groupe est minimale et pour
- 25 attribuer les signaux audio du premier groupe et leurs sources correspondantes à l'une des positions spatiales, parmi les données calculées de position spatiale du représentant du premier groupe et les données de position spatiale de la source déterminée, en fonction d'évaluations de distance d'erreur, de façon à former les deux groupes.

9. Dispositif informatique selon la revendication 8, caractérisé en ce que le module de sélection est apte à effectuer une évaluation de distance d'erreur pour un signal audio du premier groupe et sa source correspondante consistant à évaluer d'une part la distance d'erreur entre les données de position spatiale de cette source et les données calculées de position spatiale du représentant du premier groupe et d'autre part la distance d'erreur entre les données de position spatiale de cette source et les données de position spatiale de la source déterminée puis à évaluer la distance d'erreur minimale entre les deux, le module de sélection étant apte à attribuer le signal audio et sa source correspondante aux données de position spatiale de la source déterminée ou du représentant du premier groupe correspondant à la distance d'erreur minimale.

10. Dispositif informatique selon l'une des revendications 7 à 9, caractérisé en ce que les données de position spatiale de la source déterminée correspondent aux données de position spatiale du représentant du deuxième groupe.

11. Dispositif informatique selon l'une des revendications 7 à 10, caractérisé en ce que le module de sélection (120) est apte à calculer les données de position spatiale de chaque représentant de groupe en fonction de paramètres de niveau de puissance de chaque source attribuée au groupe.

12. Dispositif informatique selon les revendications 7 à 11, caractérisé en ce que le module de sélection (120) est apte à recalculer les données de position spatiale du représentant de chacun des deux groupes en déterminant une source pour laquelle la somme des distances d'erreur entre la position spatiale de cette source et celles des autres sources du groupe est minimale et le module de sélection (120) est en outre apte à ré-attribuer les sources à l'un ou l'autre des représentants d'un des deux groupes en fonction de ladite évaluation de distance d'erreur minimale.

13. Dispositif informatique selon la revendication 12, caractérisé en ce que le module de sélection (120) est apte à recalculer les données de position spatiale du représentant de

chacun des deux groupes et à ré-attribuer les sources à l'un ou l'autre des représentants d'un des deux groupes jusqu'à ce que la somme des distances d'erreur entre les représentants des deux groupes et leurs sources atteigne un minimum.

- 5     14. Dispositif informatique selon l'une des revendications 7 à 13, caractérisé en ce que le module de sélection (120) est apte à diviser un groupe jusqu'à ce qu'un nombre déterminé de groupes soit obtenu ou jusqu'à ce que la somme des distances d'erreur entre les représentants des groupes et leurs sources atteigne un seuil déterminé.
- 10    15. Procédé de traitement de signaux audio en partie pré-enregistrés correspondant chacun à une source, comprenant les étapes consistant à
- a. calculer des paramètres de niveau de puissance instantané à partir des signaux audio, les sources correspondantes étant définies par ces paramètres et par des données de position spatiale,
- 15    b. regrouper certains des signaux audio en un nombre variable de groupes de signaux audio et calculer des données de position spatiale représentatives de chaque groupe de signaux audio en fonction des données de position spatiale et des paramètres de niveau de puissance instantané de chaque source correspondante (204),
- c. traiter ces signaux audio par groupe en temps réel en fonction de données de position
- 20    spatiale représentatives du groupe (206, 208).
16. Procédé selon la revendication 15, caractérisé en ce que l'étape a. comprend en outre de sélectionner les signaux audio inaudibles en fonction des paramètres de niveau de puissance instantanés comprenant un niveau de puissance et un seuil de masquage pour chaque source
- 25    et de conserver les seuls signaux audio audibles (200).
17. Procédé selon la revendication 16, caractérisé en ce que les paramètres de niveau de puissance sont calculées pour chaque source à partir de la densité spectrale de puissance instantanée pré-calculée à partir des signaux audio en partie pré-enregistrés.



18. Procédé selon l'une des revendications 15 à 17, caractérisé en ce que l'étape c. consiste à

c1. traiter chaque groupe de signaux audio en un signal audio de pré-mixage (206),

5 c2. rassembler les signaux audio de pré-mixage afin d'obtenir un signal de mixage audible par l'auditeur (208).

19. Procédé selon la revendication 17, caractérisé en ce que l'étape c1. consiste en outre à transformer un groupe de signaux audio en un groupe de signaux vidéo texturés par utilisation d'un processeur vidéo (2020), à traiter chaque signal vidéo texturé du groupe  
10 selon des paramètres de modification du son (2022, 0224), et à rassembler et transformer les signaux du groupe en un signal audio de pré-mixage (2026).

20. Procédé selon la revendication 19, caractérisé en ce que les paramètres de modification du son comprennent un paramètre d'atténuation du son et/ou un paramètre de délai de  
15 propagation du son.

21. Procédé selon l'une des revendications précédentes, caractérisé en ce que l'étape b. consiste en outre à construire, à partir d'un premier groupe de signaux audio et de données calculées de position spatiale du groupe (2000), deux groupes de signaux audio et à calculer  
20 les données de position spatiale d'un représentant de chacun de ces deux groupes (2002 à 2012).

22. Procédé selon la revendication 21, caractérisé en ce que l'étape b. consiste en outre à déterminer, à partir du premier groupe de signaux audio, de leurs sources correspondantes  
25 et de données calculées de position spatiale du représentant du premier groupe, une source pour laquelle la somme des distances d'erreur calculées entre la position spatiale de cette source et celles des autres sources du premier groupe est minimale (2002) et à attribuer les signaux audio du premier groupe et leurs sources correspondantes à l'une des positions spatiales, parmi les données calculées de position spatiale du représentant du premier groupe

et les données de position spatiale de la source déterminée, en fonction d'évaluations de distance d'erreur, de façon à former les deux groupes (2004).

23. Procédé selon la revendication 22, caractérisé en ce que l'évaluation de distance de  
5 l'étape b. consiste pour un signal audio du premier groupe et sa source correspondante, à évaluer d'une part la distance d'erreur entre les données de position spatiale de cette source et les données calculées de position spatiale du représentant du premier groupe (A9) et d'autre part la distance d'erreur entre les données de position spatiale de cette source et les  
10 données de position spatiale de la source déterminée puis à évaluer la distance d'erreur minimale entre les deux et à attribuer le signal audio et sa source correspondante aux données de position spatiale de la source déterminée ou du représentant du premier groupe correspondant à la distance d'erreur minimale (2004).

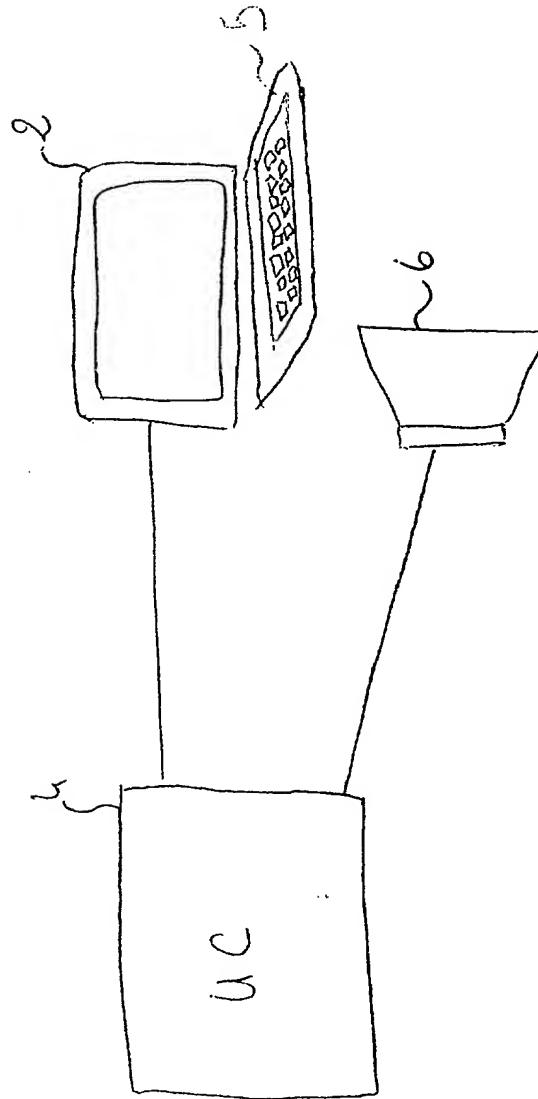
24. Procédé selon l'une des revendications 21 à 23, caractérisé en ce que les données de  
15 position spatiale de la source déterminée de l'étape b. correspondent aux données de position spatiale du représentant du deuxième groupe.

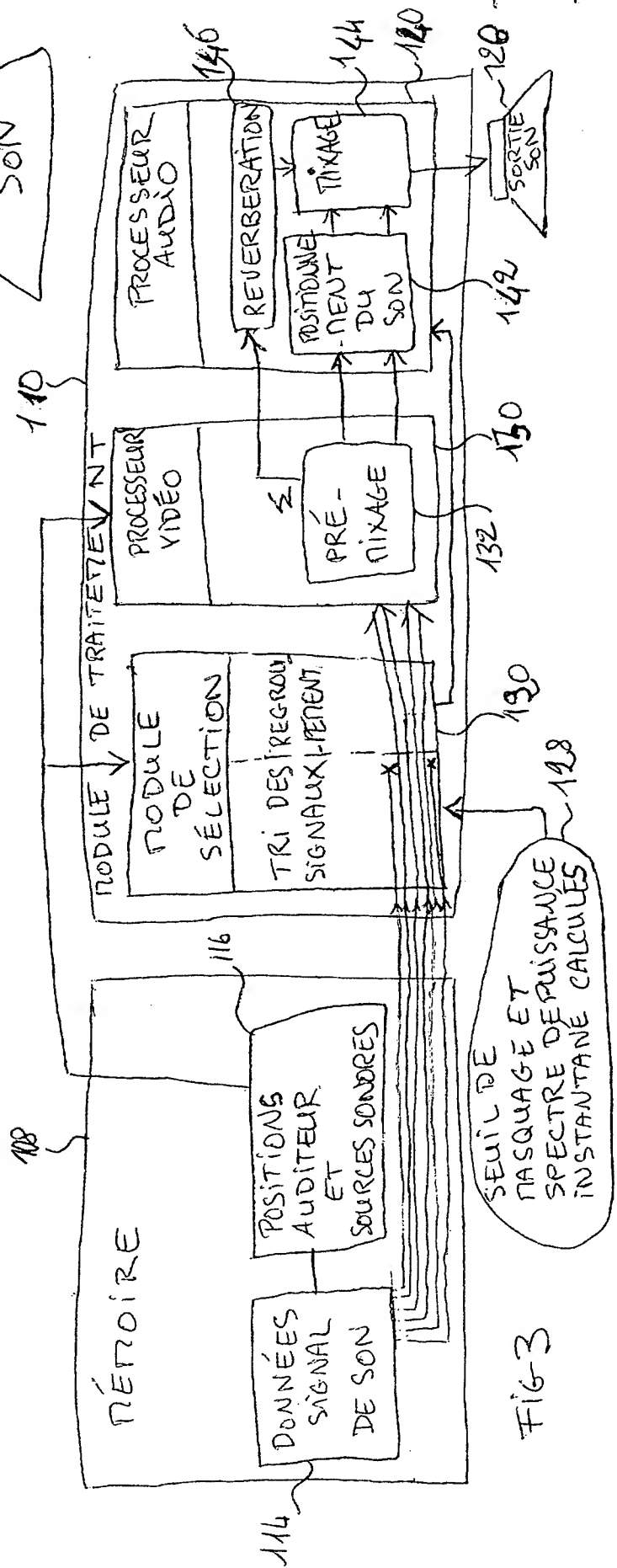
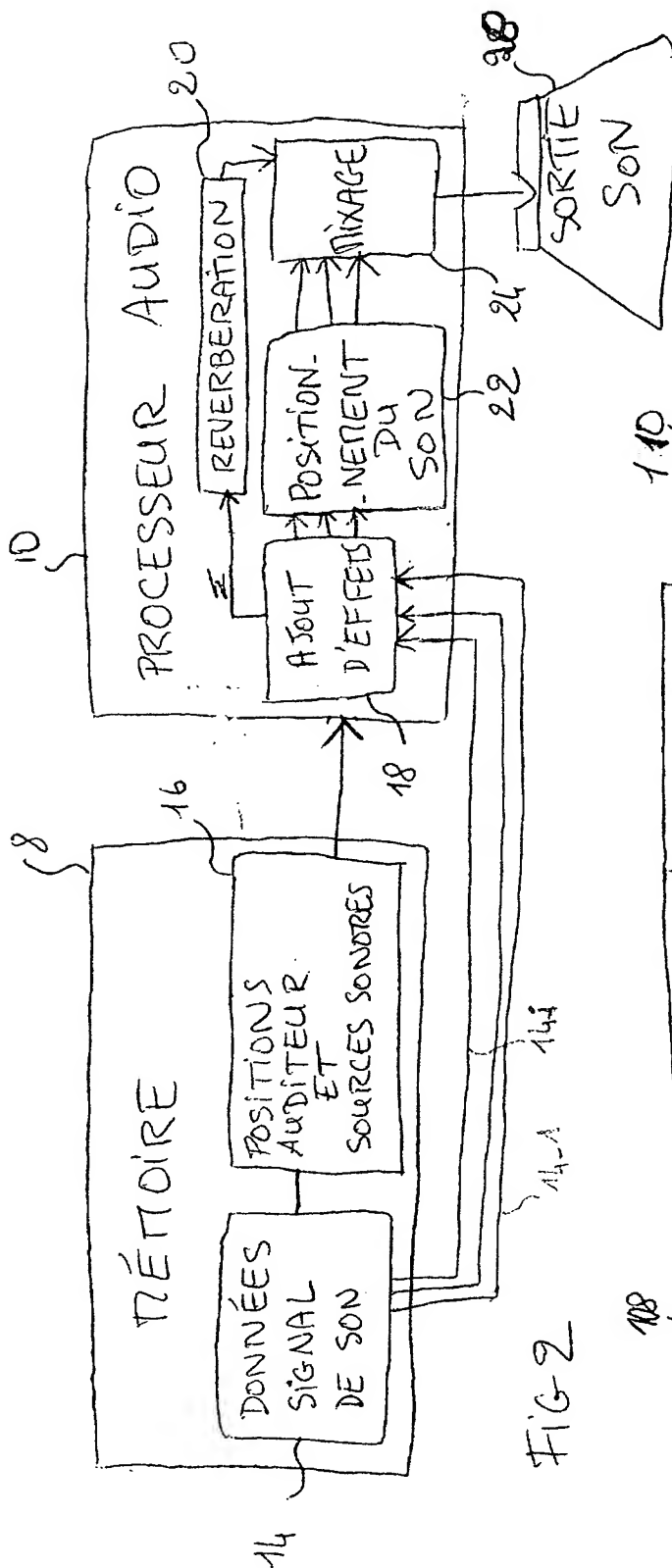
25. Procédé selon l'une des revendications 21 à 24, caractérisé en ce que l'étape b. consiste également à recalculer les données de position spatiale du représentant de chacun des deux  
20 groupes (2006) et à ré-attribuer les sources à l'un ou l'autre des représentants d'un des deux groupes (2008) jusqu'à ce que la somme des distances d'erreur entre les représentants des deux groupes et leurs sources atteigne un minimum (2010).

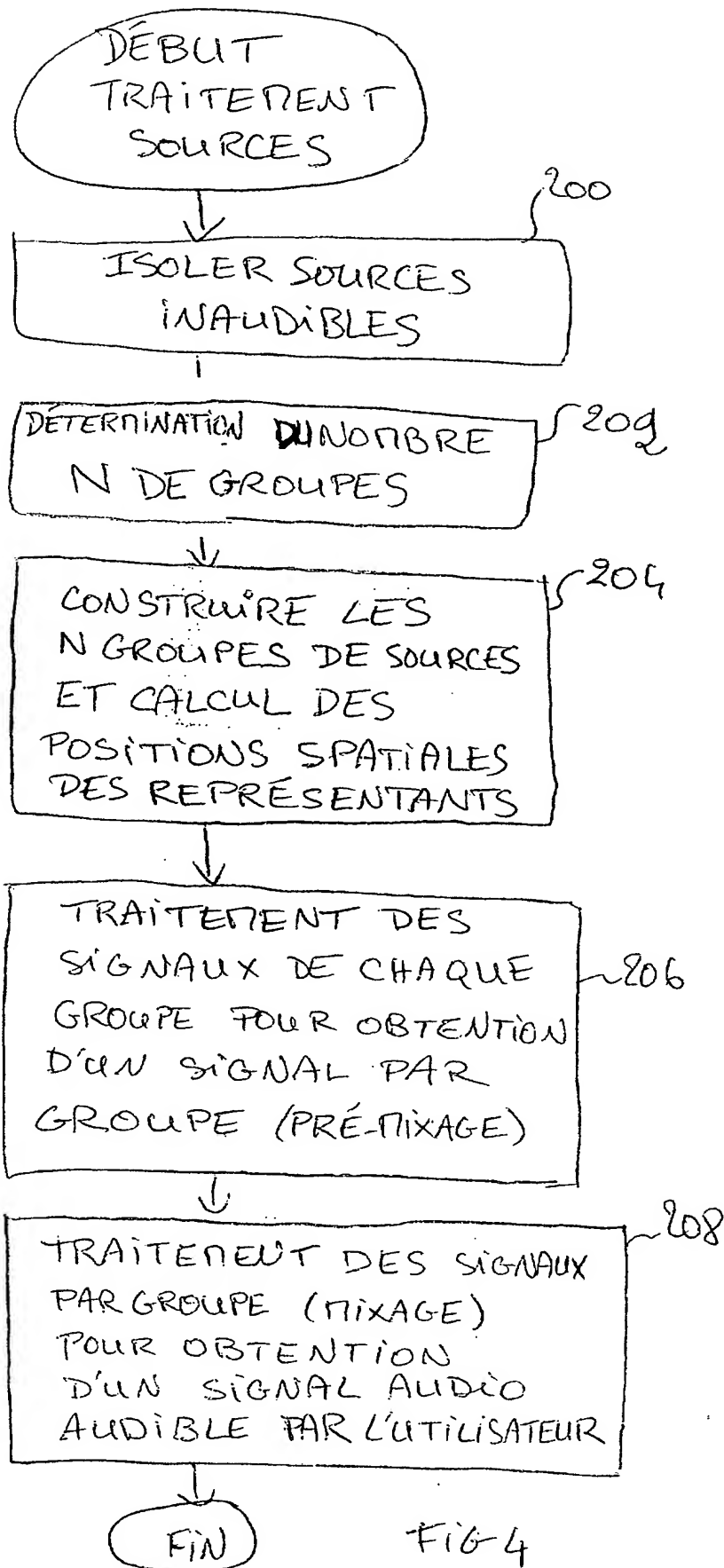
26. Procédé selon les revendications 21 à 25, caractérisé en ce que l'étape b. consiste à  
25 diviser un groupe jusqu'à ce qu'un nombre déterminé de groupes soit obtenu ou jusqu'à ce que la somme des distances d'erreur entre les représentants des groupes et leurs sources atteigne un seuil déterminé (2012).

1/7

FIG 1







H17

# PROCÉDÉ DE DIVISION PAR GROUPES - 204

DÉFINITION DU  
1<sup>er</sup> GROUPE - 2000  
CALCUL POSITION  
SPATIALE DU REPRÉ-  
-SENTANT C1

CHOIX SOURCE  $S_i$  DU GROUPE  
VÉRIFIANT  $\Pi_{in} [E_n = \sum_k d(S_i, S_k)]$  2002  
 $C2 \leftarrow S_i$

DIVISION DU GROUPE EN 2:  
ATTRIBUTION DES SOURCES  
À C1 OU À C2 REPRÉSENTANTS  
DES 2 GROUPES 2004

RECALCUL POSITION SPATIALE  
DES REPRÉSENTANTS DES  
GROUPES : C1 ET C2 2006

RÉATTRIBUTION DES  
SOURCES À C1 OU À C2 2008

NON

$$E_n + E_n = \Pi_{in local}$$

2010

oui

NON

OBTENTION DE N GROUPES?  
ou  $\sum E_n < \text{seuil}$  2012

FIG. A

oui

FIN

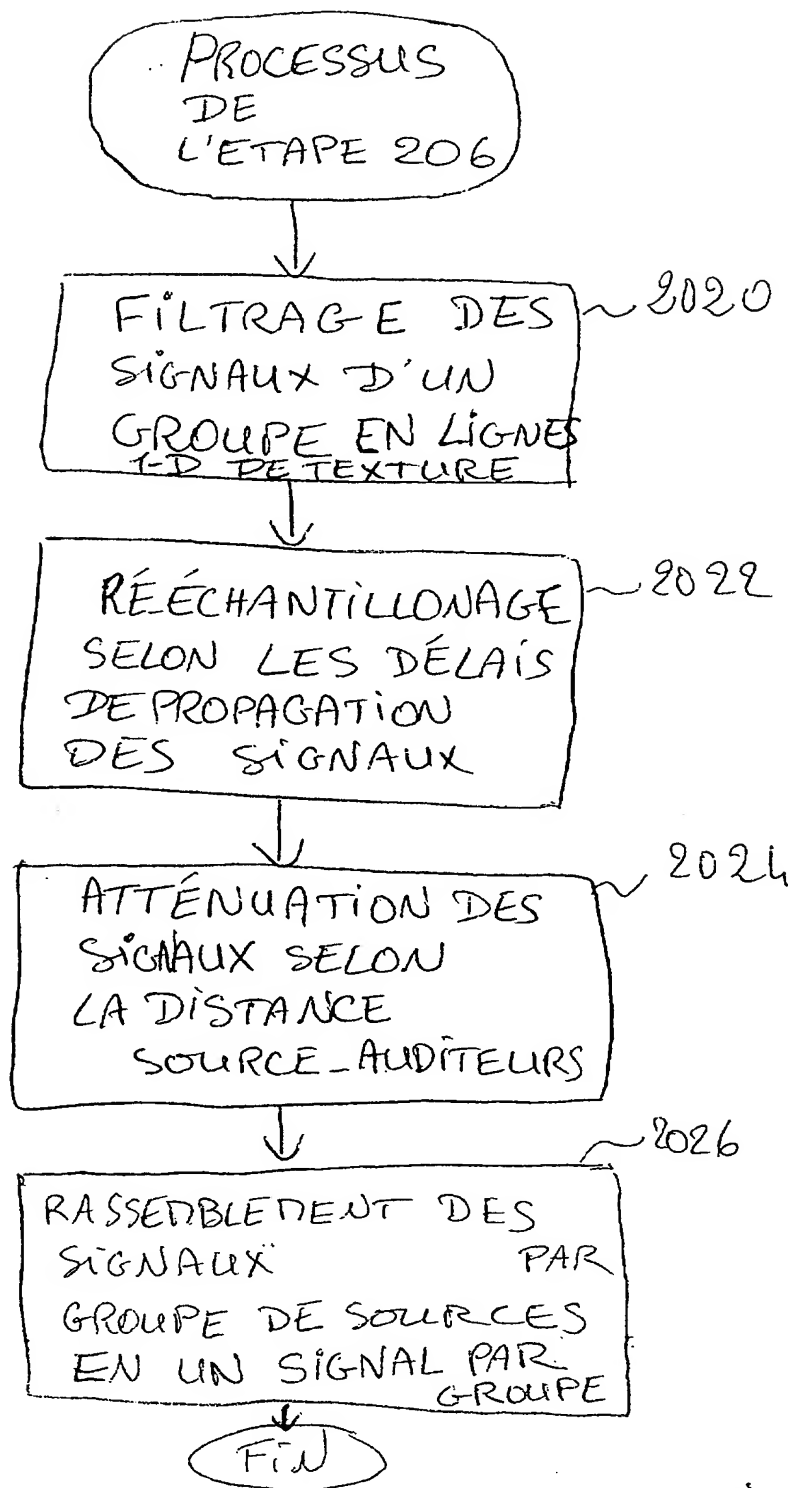
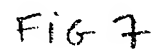
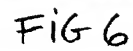


FIG. 4B.





7/7

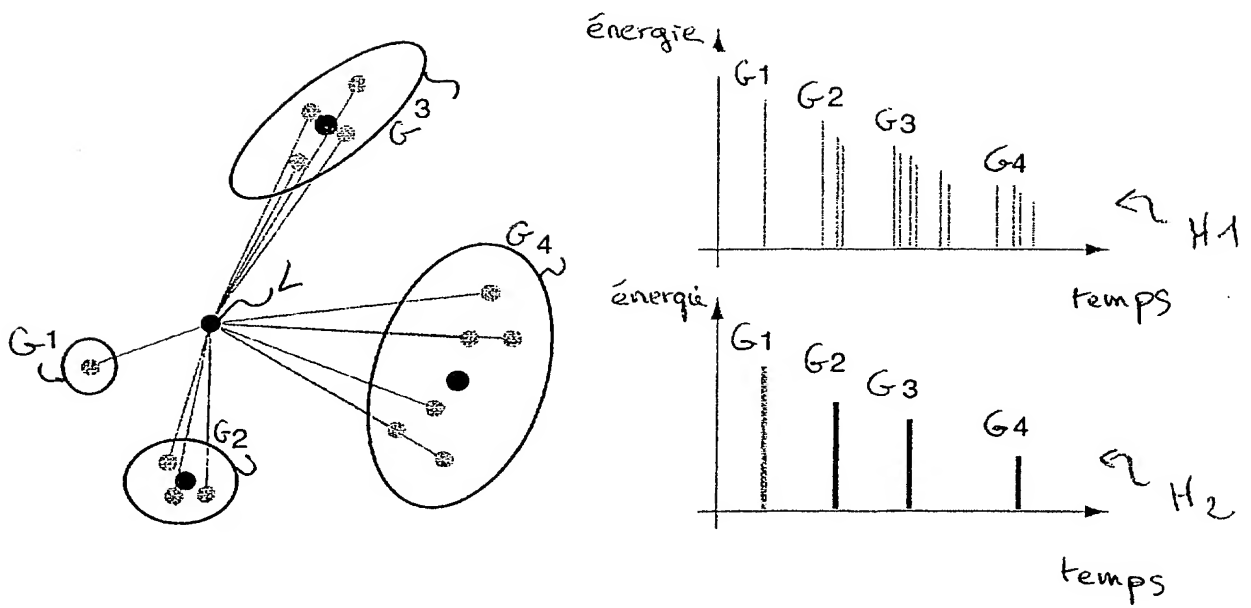


FIG 8



# BREVET D'INVENTION

## CERTIFICAT D'UTILITÉ

Code de la propriété intellectuelle - Livre VI



N° 11235\*02

## DÉPARTEMENT DES BREVETS

26 bis, rue de Saint Pétersbourg  
75800 Paris Cedex 08

Téléphone : 01 53 04 53 04 Télécopie : 01 42 93 59 30

DÉSIGNATION D'INVENTEUR(S) Page N° 1.. / 1..

(Si le demandeur n'est pas l'inventeur ou l'unique inventeur)

Cet imprimé est à remplir lisiblement à l'encre noire

DB 113 W / 260899

|                                                                                                                                                                                                                            |                      |                               |         |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|-------------------------------|---------|
| <b>Vos références pour ce dossier</b><br>(facultatif)                                                                                                                                                                      |                      | INRIA 72 (120905)             |         |
| <b>N° D'ENREGISTREMENT NATIONAL</b>                                                                                                                                                                                        |                      | 03 13875 du 26 novembre 2003  |         |
| <b>TITRE DE L'INVENTION</b> (200 caractères ou espaces maximum)<br>Dispositif et méthode perfectionnés de spatialisation du son                                                                                            |                      |                               |         |
| <b>LE(S) DEMANDEUR(S) :</b><br>INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE                                                                                                                            |                      |                               |         |
| <b>DESIGNE(NT) EN TANT QU'INVENTEUR(S) :</b> (Indiquez en haut à droite «Page N° 1/1» S'il y a plus de trois inventeurs, utilisez un formulaire identique et numérotez chaque page en indiquant le nombre total de pages). |                      |                               |         |
| Nom                                                                                                                                                                                                                        |                      | TSINGOS                       |         |
| Prénoms                                                                                                                                                                                                                    |                      | Nicolas                       |         |
| Adresse                                                                                                                                                                                                                    | Rue                  | 505 Chemin de Rabiac Estagnol |         |
|                                                                                                                                                                                                                            | Code postal et ville | 06600                         | ANTIBES |
| Société d'appartenance (facultatif)                                                                                                                                                                                        |                      |                               |         |
| Nom                                                                                                                                                                                                                        |                      | GALLO                         |         |
| Prénoms                                                                                                                                                                                                                    |                      | Emmanuel                      |         |
| Adresse                                                                                                                                                                                                                    | Rue                  | 14 rue Miron                  |         |
|                                                                                                                                                                                                                            | Code postal et ville | 06000                         | NICE    |
| Société d'appartenance (facultatif)                                                                                                                                                                                        |                      |                               |         |
| Nom                                                                                                                                                                                                                        |                      | DRETTAKIS                     |         |
| Prénoms                                                                                                                                                                                                                    |                      | George                        |         |
| Adresse                                                                                                                                                                                                                    | Rue                  | 21 avenue Henri Barbusse      |         |
|                                                                                                                                                                                                                            | Code postal et ville | 06000                         | NICE    |
| Société d'appartenance (facultatif)                                                                                                                                                                                        |                      |                               |         |
| <b>DATE ET SIGNATURE(S)</b><br><b>DU (DES) DEMANDEUR(S)</b><br><b>OU DU MANDATAIRE</b><br>(Nom et qualité du signataire)<br>Paris, le 30 décembre 2003<br>Jean-Yves PLAÇAIS<br>92-1197 (B)(M)                              |                      |                               |         |